MULTIMEDIA UNIVERSITY

STUDENT ID NO

TABLE NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2019/2020

TDS2201 – STATISTICAL DATA ANALYSIS

(All sections / Groups)

7 MARCH 2020
2:30 pm – 4:30 pm
(2 Hours)

Question	Marks
1	
2	
3	
4	
Total	

INSTRUCTION TO STUDENT

1. This question paper consists of 9 printed pages (inclusive of the front page) with 4 questions only. A formula sheet is attached (page 9).
2. The distribution of marks for each question is given. Attempt **ALL FOUR** questions. Show ALL of your working steps clearly.
3. Please write your **student ID** and **table no** in the space at the top right corner on this front page.
4. Please write all your answers in the space provided for each question in this question paper.

Question 1 (10 Marks)

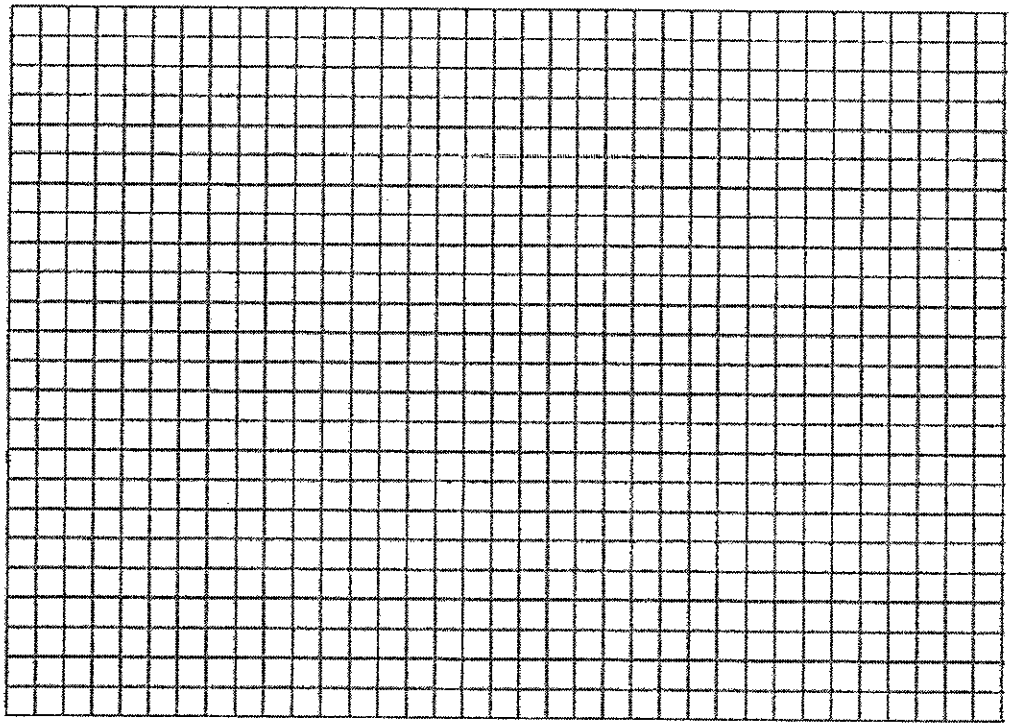
a) Given is sodium content (mg) per 2 tablespoon served in 10 different peanut butters:

120, 85, 140, 80, 150, 175, 335, 110, 220, 450

i) Determine the empirical quantile and the Normal distribution theoretical quantile of the data. [2 marks]

i	Probability (p_i)	Theoretical quantile (π_{p_i})	Empirical quantile
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

ii) Sketch a quantile-quantile plot for the data at the graph paper given below. Make sure your graph is readable. Label your axis. [2 marks]



... Continued

iii) Is the data normally distributed? Explain your answer. [2 marks]

v) Comment on the shape of distribution of data. [1 mark]

iv) Construct a stem-and-leaf graph to verify your answer in v). Use unit of 100 in stems and truncate leaves to unit of 10. [1 mark]

b) The defect of an automobile part in certain model is a random variable with mean 5.5 and standard deviation 4.1. Among 50 randomly selected cars of this model, how likely is that the sample average of major defects exceeds 6? [2 marks]

... Continued

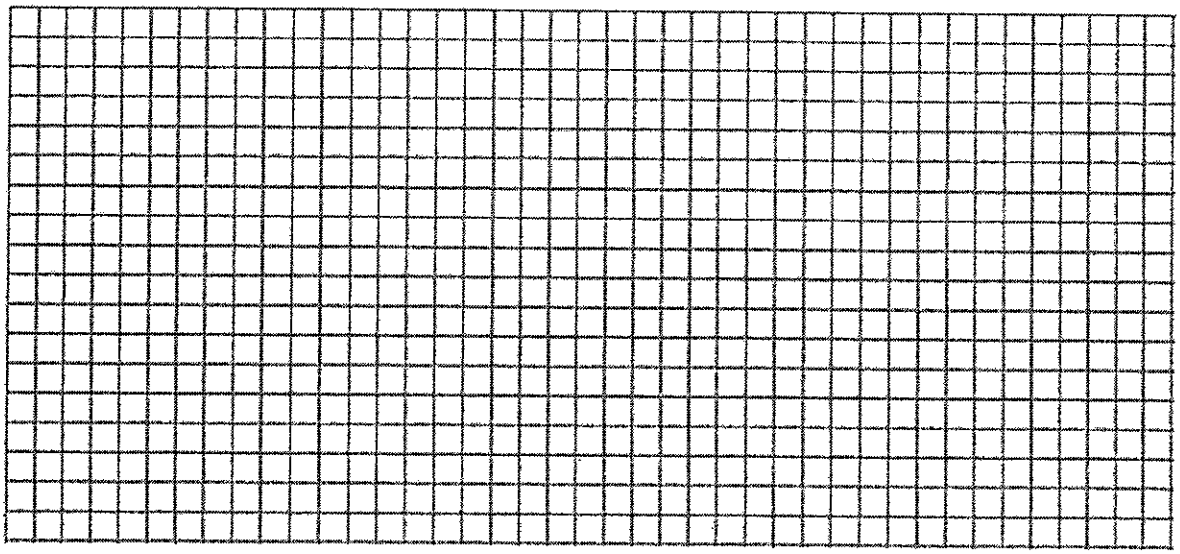
Question 2 (10 Marks)

Exposure of microbial products especially endotoxin may have an impact on vulnerability of allergic diseases. The following data on endotoxin concentration (EU/mg) in settle dust for one sample of urban home and another of farm home are given as follows:

Urban:	6,	5,	11,	33,	4,	5,	80,	18,	35,	17
Farm:	4,	14,	11,	9,	9,	8,	4,	17,	5,	8,
	11,	9,	3,	2,	1					

a) For each sample, find the mean, median, quartiles, interquartile range and list all the outliers if exists. [6 marks]

b) Sketch the comparative boxplot for the above data. Draw your boxplot horizontally and label your axis. [2 marks]



... Continued

c) Comments on the similarities and differences shown on the comparative boxplots of endotoxin concentrations on urban and farm homes. [1 mark]

d) From a) to c) explain why the sample mean of endotoxin concentration of urban home so much different from its median? [1 mark]

... Continued

Question 3 (10 Marks)

Table below is the estimated repair costs from two workshops A and B on 9 defective items.

Item	1	2	3	4	5	6	7	8	9
Workshop A	17.6	20.2	59.5	71.3	83.0	26.3	45.7	12.2	18.5
Workshop B	17.3	19.2	61.5	74.5	81.2	24.7	43.2	11.3	19.1

Assume that each of the estimated cost is approximately normally distributed.

a) Construct a 95% confidence interval to determine any significant differences in their repair costs. Justify your answer. [5.5 marks]

... Continued

b) The standard deviation on the difference in repair cost between workshops A and B is $\sqrt{2}$. Based on the above data, do you think that $\sigma > \sqrt{2}$? State and test the relevant hypothesis using 0.05 level of significance. [4.5 marks]

... Continued

Question 4 (10 marks)

Given is the data on the height and weight of 10 randomly selected students in Multimedia University

Height (cm)	170	150	140	175	180	155	148	160	173	163
Weight (kg)	70	60	45	90	92	51	48	65	80	74

a) The weight and height of students are assumed to be linearly related. Fit the regression model for those two variables. [2 marks]

b) What happened to the weight if the height of a student is increased by 1 cm? [1 mark]

c) Estimate the average weight when the average height is 155 cm. Calculate the residual of weight at this average height. [2 marks]

d) How much variability of weight can be explained by average height? Write down the notation and name of statistic you used to answer this question. [2 marks]

e) What is the assumption on residual when we fit a linear regression model? What plot can be used to check this assumption? Under what condition of the plot we conclude that the assumption is not violated? [3 marks]

... Continued

Formula Sheet

Confidence Interval

1) $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

2) $\bar{x} \pm t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}}$

3) $\bar{x}_1 - \bar{x}_2 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

4) $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}(n_1 + n_2 - 2) \times s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

5) $\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}(v) \times \sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}, \quad v = \frac{(w_1 + w_2)^2}{\frac{w_1^2}{n_1 - 1} + \frac{w_2^2}{n_2 - 1}}, \quad w_i = \frac{s_i^2}{n_i}$

6) $\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

7) $\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

8) $\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}$

9) $\frac{s_1^2}{s_2^2} \frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_2, v_1)$

Test of Significance

1) $\text{Test statistics} = \frac{\text{sample estimate - null value}}{\text{standard deviation of sample estimate}}$

2) $X^2 = \sum\sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

End of Page.